Multimodal Interaction Design for Public Window Displays: A Case Study

Xiao-Nan Liu
School of Arts and Media
Tongji University
Shanghai, China
xiaonanliu@tongji.edu.cn

Min-Zhi Shao
School of Arts and Media
Tongji University
Shanghai, China
shao@tongji.edu.cn

CHI 2020 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
DOI: https://doi.org/10.1145/3334480.3382838

Abstract
We present a case study of multimodal interaction design for public window displays. Using a classic fairytale as the theme story, a prototype system that integrates mobile, gesture, tangible, touchscreen, and puppet interfaces has been implemented. The preliminary field deployment results demonstrate that our interactive window is well-received, with a significantly extended duration of user's interaction time. We conclude with a discussion of lessons learned and potential new research problems for interactive public window design. We believe our findings are useful in future design for interactive shop windows, theater showcases, and exhibition displays.

Author Keywords
Public displays; interactive window; multimodal interfaces; mental models; interactive storytelling.

CSS Concepts
• **Human-centered computing~Interaction design**; *Interaction design process and methods*; User interface design.

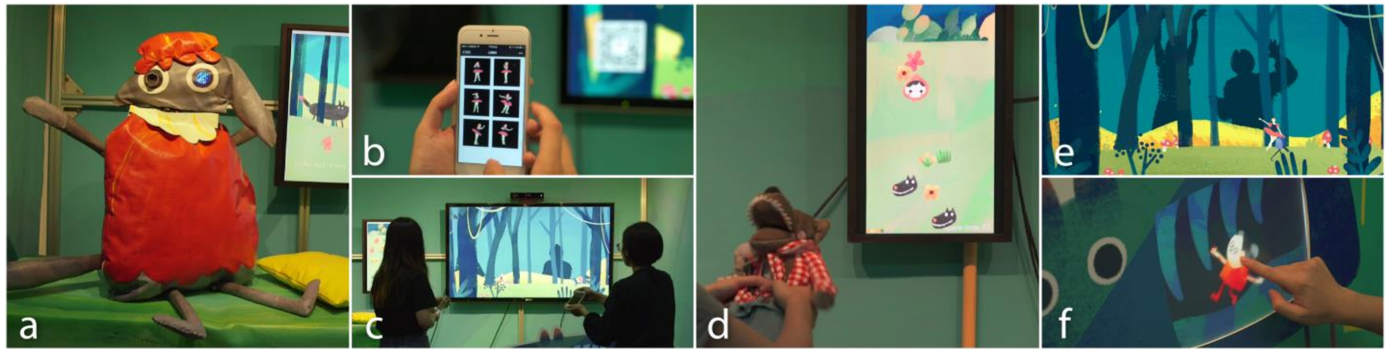

Figure 1: Interaction modalities: a) puppet; b,c) mobile; d) tangible; e) gesture; f) touchscreen.

Figure 2: Display layout.

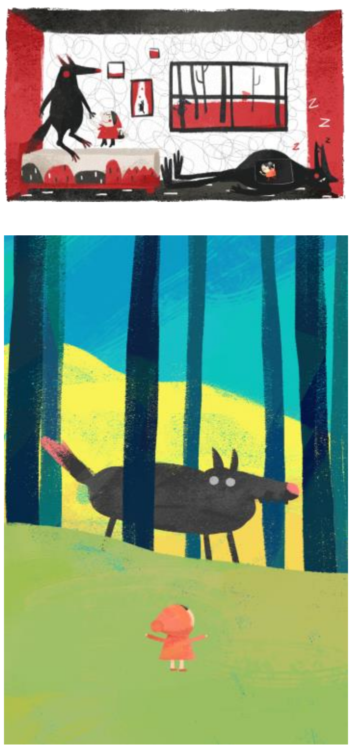

Figure 3: Early concept sketches.

Introduction

With the rapid development of interaction technologies, public displays have become increasingly popular in our daily lives. Research in recent years has made significant progress in understanding the design and user experiences for this new type of public engagement [8,1], e.g., using silhouettes or mirror images [9,11], dynamically transparent window [3], animatronic motion [6], and adaptive interfaces [12], to arouse curiosity in passers-by and encourage them to explore the display content. A common problem, however, is that passers-by usually pay little attention to them [4]. In many cases, even if the display can successfully attract the attention of passers-by and make them notice the interactivity, they often refuse to participate due to lack of interest in the content, or reluctance to spend time and effort to learn the interaction interfaces.

To encourage user participation, the motivation to interaction is critical. The interactive display in the public environment should make full use of its presentation to attract the attention of passers-by, inform them of the interactivity, so that they can actively engage, understand, and absorb the content.

In this work, we aim to design such a system: 1) has a unified narrative theme that stimulates people's curiosity through storytelling and fun playing, to engage them in participation and interaction;
2) connects real and virtual worlds by staging the story on an open and three-dimensional display, and using physical props to create an immersive interaction experience; 3) supports multimodal and multi-user interaction, offering more choices and experiences for participants of different ages and diverse interests;
4) adopts mental models, experimenting with people's intuitive perception for public displays. To this end, we present a multimodal interactive window prototype that integrates mobile, gesture, tangible, touchscreen, and puppet interfaces. We outline our design rationale and implementation process, report the field deployment findings, and conclude with a discussion of lessons learned and potential new research problems for interactive public window design.

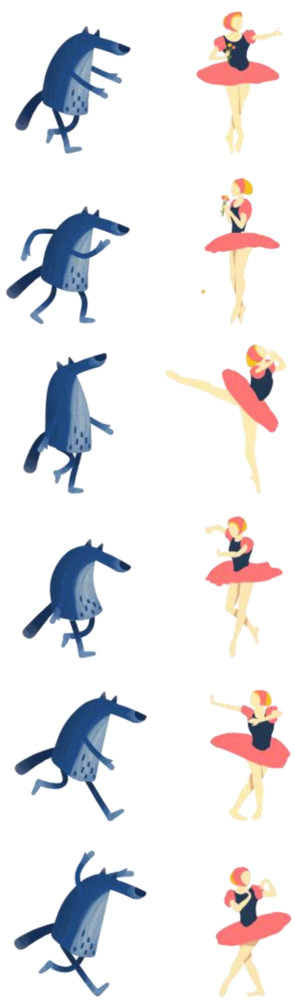

Figure 4: Six different animations for each of Little Red and Big Wolf.

Design

We study and evaluate the design rationale from four perspectives: narrative theme, display layout, interaction modalities, and mental models.

Narrative theme

The first problem, when designing a multimodal interactive display, is how to unify multiple interaction modules to a single narrative theme in a natural and coherent way, so that each module is structurally independent and at the same time logically related to each other. As a prototype experiment, a well-known story will probably help the user quickly understand and familiar with the multimodal interaction interfaces. A classic fairytale may further inspire fantasy and exploration. We selected 'Little Red Riding Hood' mainly for these purposes. Four plot scenes were chosen from the story, namely, Little Red picking flowers in the forest, Big Wolf chasing Little Red, Big Wolf swallowing up Grandma, and Big Wolf disguising as Grandma, to correspond to different interaction modalities.

Display layout

We chose Grandma's room to be the main story scene, and set the other scenes as the room's window view and framed picture, thus place the four plot scenes in one unified display layout. The display's dimensions were set to 300x170x150cm (with the aspect ratio of the façade approximately 16:9). Instead of a conventional closed window display (e.g., shop window), we decided for an open display to make it more inviting and friendly. A bright and colorful palette was applied to lighten up the overall tone. In particular, we set the interactable characters in each plot scene to red color (e.g., Little Red in the forest, Grandma inside Big Wolf's belly, Big Wolf in bed), giving users a hint of the interaction logic and structure; see Figure 2. Two early concept sketches are shown in Figure 3.

Interaction modalities

Corresponding to four plot scenes, we identified five interaction modalities: mobile, gesture, tangible, touchscreen, and puppet. The mobile module and the gesture module exhibit Big Wolf chasing Little Red in the forest scene; the tangible module tells the story of Little Red picking flowers in the forest; the touchscreen module displays Big Wolf swallowing up Grandma; the puppet module presents Big Wolf dressed as Grandma, waiting for Little Red to come home; see Figure 1.

Mental models

Understanding how users intuitively perceive the world around a display is essential for the interaction design. Müller et al. [8] have analyzed existing public display applications and identified four prevailing mental models based on metaphors from the real world: poster, window, mirror, and overlay. We decided to experiment with all four mental models in a single unified design. To our knowledge, this may be the first attempt for public displays.

Implementation

Mobile

A 55" LED display screen is used as the window in Grandma's room. Shown outside the window is the introductory part of the story, Little Red picking flowers in the forest. Two users can use their mobile phones to scan the QR codes located on the left and right corners of the screen, which correspond to Little Red and Big Wolf, respectively. The host computer reads the built-in accelerometer output from the mobile phones using an HTML5 program linked to the QR codes. This enables

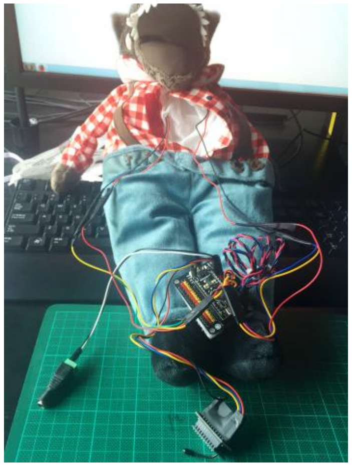

Figure 5: Electronic Big Wolf (made of an off-the-shelf IKEA plush doll).

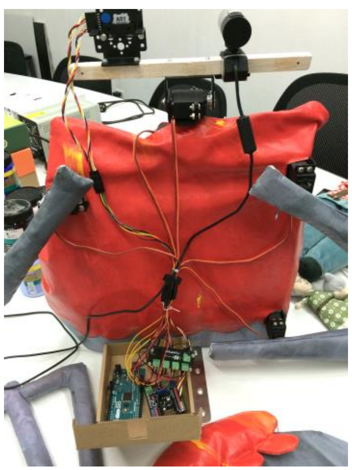

Figure 6: Back view of Grandma Wolf.

the users to control the character movement by simply tilting their mobile phones; see Figure 1c. Six different preprogrammed animations can be selected for each character, e.g., Little Red giving directions, calling for help, strewing flowers, and so on; see Figure 1b and Figure 4.

In the context of mental model theory, the room's window view design falls into the category of the window metaphor, which as aptly described in [8], tends to create the illusion of a link to a remote, often virtual, location. Mobile phones, with various built-in sensors and a familiar interface, are ideal as ubiquitous interaction devices to connect with large and often out-of-reach displays [2].

Gesture

At night, 'monster' silhouettes can be seen in the depths of the forest. These silhouettes are the passers-by's profiles captured in real time by a Microsoft Kinect mounted at the top of display screen; see Figure 1e. The large screen makes it easy for passers-by to discover their own gestures, understand the display's interactivity, and participate in it.

In gesture silhouette, we experiment with the mirror metaphor in mental model [8]. As previous research has repeatedly demonstrated, making passers-by a part of the public display is one of the most effective ways to attract and retain their attention [9,11].

Tangible

Next to Grandma's bed, a picture frame is hung on the wall with a 24" flat screen monitor embedded behind it. The user can take on the role of Big Wolf character by manipulating a plush doll (placed inside the open display window) and interact with Little Red in the picture; see Figure 1d. Embedded in the plush doll is a set of accelerometer, flex, and pressure sensors connected to an Arduino Nano microcontroller; see Figure 5. The host computer collects the sensor output wirelessly through an XBee module and transmits the data to a Flash animation program to control the movement of the virtual Little Red character on display.

The framed picture is a design based on the poster metaphor in mental model, which as observed in [8], often tends to be ignored by passers-by due to the association with traditional advertising posters. We try to counter this preconception by connecting the virtual display with an interaction-ready physical prop. Interesting research in interactive cartoon [5] and interactive theatrical performance [7] has shown that using plush dolls in a physical/virtual hybrid design attracts the attention of users and audiences, and makes their interaction experiences immersive and compelling.

Touchscreen

A cardboard cutout of Big Wolf, whose open belly is embedded with a touchscreen, stands at the room's lower right corner; see Figure 1f. This part illustrates Grandma being swallowed by Big Wolf, is hitting his belly with all her arms and legs, hoping someone come to pull her out. A touch-driven elastic deformation algorithm is implemented for this module, which is intuitive to understand and simple to interact with. This is particularly suitable for children. So we position it at a height that is easy for them to reach. At the same time, children will not block the adult users standing behind interacting with other modules, e.g., gesture and puppet.

Figure 7: Preprogrammed 3D puppet animations.

Figure 8: Face tracking test in puppet interaction.

Overlay metaphor in mental model refers to displaying content within another context, and is typically created by image projection [8]. In our design, it is realized by a capacitive touchscreen embedded within a cardboard cutout, affording accurate touch interaction. One particular problem for touch interaction in multimodal public displays is finding efficient affordance; the conventional design is standard button and call-to-action [9]. Our solution is adding a random force in the elastic deformation algorithm when the touchscreen is idle (i.e., not touched), so that Grandma is always in (trembling) motion.

Puppet

This part corresponds to the end of the story, Big Wolf disguising as Grandma, sits in bed waiting for Little Red to come home. Grandma Wolf is a physical puppet; see Figure 1a and Figure 6. He reacts to the user's action with a set of preprogrammed 3D animations, e.g., waving when she approaches, coquetting when she leaves; see Figure 7. He also follows the user's movement by turning his head and gazing at her face with a Microsoft LifeCam embedded in his right eye. Embedded in his left eye is a small LED display, which plays bitmap animation when face tracking is on. The face tracking algorithm is programmed using OpenCV (Figure 8), and the puppet's animation is driven by five servo motors controlled by an Arduino Mega microcontroller.

Our track-and-react puppet is designed for the purpose of better indicating availability and encouraging engagement. Research has found the animatronic motion significantly attracts more looks and touches than on-screen display [6], and dynamic tracking of people's motion raises their curiosity and desire for exploration [3].

The prototype system runs under Windows 10 Pro on a Xeon E5620 processor (2.40GHz, 12GB), with graphics rendered on an NVIDIA GTX1070 card. The software is written with C++, OpenCV, Arduino, Flash, and HTML5.

Evaluation

We conducted a preliminary field deployment for our prototype system in two locations. The first was on the ground floor in the historic Shanghai 1933 building for three days, and the second was in the gallery of Shanghai Grand Theatre for two days during a weekend; see Figure 9.

Despite some minor technical glitches during the first two days at Shanghai 1933, overall the statistics are surprisingly consistent for all five days. The vast majority of people passing by the installation at least glanced at it. We thus count only those who stopped as the interaction users. For the third day, we logged 3,041 passers-by, of which 895 (29.4%) interacted. Among them, 18 (2.0%), 488 (54.5%), 15 (1.7%), 75 (8.4%), and 146 (16.3%) persons tried mobile, gesture, tangible, touchscreen, and puppet interactions, respectively. We found the overall response received from users very positive. The mean interaction duration was 351s (median = 127s) with a standard deviation of 627s. Interacting with our prototype system resulted in a significantly extended duration comparing to previously reported research, e.g., in [4,3,11].

The following is our main findings based on recorded video, in-situ observations, and interviews: 1) Mirrored user silhouettes (in gesture interaction) were the most

Figure 9: Field deployment at Shanghai 1933 and Shanghai Grand Theatre.

effective in informing interactivity. The face tracking puppet attracted the second largest number of user interactions. 2) The touchscreen tended to lead long interaction time for children, who also liked to make up stories while playing. It was evident that touchable virtual characters with the effects of deformation or animation would stimulate children's fantasy and creativity. 3) Many users performed various creative acts and struck poses in gesture interaction. While the average duration was relatively short, we noticed that it was very effective in conveying interaction affordance. When users noticed one interaction modality, they would often attempt to discover and explore other ones. 4) Most users did not seem to pay much attention to the QR codes on the screen, let alone scan them to interact. We tried animated QR codes in our second field deployment, but saw little improvement. Perhaps people thought these codes were for commercial purposes or had other security concerns. 5) For tangible interaction (plush doll), most users did not notice its interactivity, probably because of social etiquette that says touching in public is impolite. We consider attaching a vibration motor and LED lights to the plush doll to better indicate affordance. 6) The face tracking puppet caught the fascination of many users. They would stop to watch, wave, and talk to him in amazement. We plan to place a Leap Motion controller inside puppet's mouth to capture the hand gestures, and a microphone to detect the voices. 7) Between children and adults, the biggest difference we noticed was children liked the touchscreen while adults the gesture, and they both liked the puppet. We believe the multimodal choices and layout make it possible for multiple people (on many occasions, children and parents) to play and have fun at the same time.

Future Work

The user transitions between modalities are inherent in multimodal interaction. A clear understanding of how users transition between modalities (e.g., the order of modality usages, the conditions of modality transitions, the multiple and repeated modality usages) is essential for real-world applications [9]. We plan to use the conversion diagrams for a quantitative and in-depth evaluation of our design.

One interesting idea that is worth further exploring is the design of 'connecting devices' to facilitate user transitions between interaction modalities, in particular, between real and virtual worlds [10].

Multimodal storytelling is inherently not sequential. So another perspective to approach the problem is studying and developing the principles and language of parallel storytelling, and applying them to the design of multimodal interfaces.

On a side note, our display as a whole does not fit into any of the four established mental models, which are two-dimensional and flat. We think stage (as in theater stage) might be a suitable metaphor, and there are principles, techniques, and tools in theater stage design to be learned for interactive storytelling and window displays.

Acknowledgements

We thank Xiang-Zhen Fan for her work in the field study, the anonymous reviewers for their comments and suggestions, and Jiao-Lei Li at Shanghai 1933 and Jue Wang at Shanghai Grand Theatre for their help with our field deployment.

References

[1] Florian Alt, Stefan Schneegaß, Albrecht Schmidt, Jörg Müller, and Nemanja Memarovic. 2012. How to Evaluate Public Displays. In *Proceedings of the International Symposium on Pervasive Displays* (PerDis '12). https://doi.org/10.1145/2307798.2307815

[2] Rafael Ballagas, Jan Borchers, Michael Rohs, and Jennifer G. Sheridan. 2006. The Smart Phone: A Ubiquitous Input Device. *IEEE Pervasive Computing*, 5, 1 (January 2006), 70-77. https://doi.org/10.1109/MPRV.2006.18

[3] Peter Dalsgaard and Kim Halskov. 2009. Dynamically Transparent Window. In *Extended Abstracts on Human Factors in Computing Systems* (CHI EA '09), 3019-3034. https://doi.org/10.1145/1520340.1520429

[4] Elaine M. Huang, Anna Koster, and Jan Borchers. 2008. Overcoming Assumptions and Uncovering Practices: When Does the Public Really Look at Public Displays? In *Proceedings of the International Conference on Pervasive Computing* (Pervasive '08), 228-243. https://doi.org/10.1007/9783540.79576614

[5] Michael Patrick Johnson, Andrew Wilson, Bruce Blumberg, Christopher Kline, and Aaron Bobick. 1999. Sympathetic Interfaces: Using a Plush Toy to Direct Synthetic Characters. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '99), 152-158. https://doi.org/10.1145/302979.303028

[6] Wendy Ju and David Sirkin. 2010. Animate Objects: How Physical Motion Encourages Public Interaction. In *Persuasive Technology*, Springer Berlin Heidelberg, 40-51. https://doi.org/10.1007/9783642.1322616

[7] Ali Mazalek, Michael Nitsche, Claudia Rébola, Paul Clifton, Andy Wu, Nick Poirier, and Firaz Peer. 2012. Pictures at an Exhibition: Design of a Hybrid Puppetry Performance Piece. In *Proceedings of the International Conference on Entertainment Computing* (ICEC '12), 130-143. https://doi.org/10.1007/9783642.33542612

[8] Jörg Müller, Florian Alt, Daniel Michelis, and Albrecht Schmidt. 2010. Requirements and Design Space for Interactive Public Displays. In *Proceedings of the ACM International Conference on Multimedia* (MM '10), 1285-1294. https://doi.org/10.1145/1873951.1874203

[9] Jörg Müller, Gilles Bailly, Thor Bossuyt, and Niklas Hillgren. 2014. MirrorTouch: Combining Touch and Mid-air Gestures for Public Displays. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices & Services* (MobileHCI '14), 319-328. https://doi.org/10.1145/2628363.2628379

[10] Honoka Ozaki, Yasushi Matoba, and Itiro Siio. 2018. Can I GetToyIn?: A Box Interface Connecting Real and Virtual Worlds. In *Extended Abstracts on Human Factors in Computing Systems* (CHI EA '18). https://doi.org/10.1145/3170427.3186472

[11] Lida Theodorou and Patrick G.T. Healey. 2015. 12 Way Mirror: Reflecting on Window Displays. In *Extended Abstracts on Human Factors in Computing Systems* (CHI EA '15), 647-654. https://doi.org/10.1145/2702613.2702955

[12] Martin Tomitsch, Christopher Ackad, Oliver Dawson, Luke Hespanhol, and Judy Kay. 2014. Who Cares about the Content? An Analysis of Playful Behaviour at a Public Display. In *Proceedings of the International Symposium on Pervasive Displays* (PerDis '14). https://doi.org/10.1145/2611009.2611016